

CLAIMS

1. Method for performing automatic analyses and comparisons of patents and technical descriptions of engineering systems, based on classifying functions to associated subsystems and sub-functions as well as functional elements to associated physical components, organizing such data in different forms according to the scope of the analysis, characterized in that all system components described in an examined text are identified, ordered and classified as a hierarchy in terms of detail/abstraction level and, further, in term of categories like "assembly", "part", "portion", all functional links existing between said identified components of the examined system being recognized so that all secondary products and, among these, a main product of the examined system are identified.
2. Method according to Claim 1, characterized in that the identification of all system components described in the examined text is performed according to the following procedure:
 - a. searching for numeric characters in a text;
 - b. for each number, taking into account a range of preceding and following words, each range constituting a row of the matrix of candidate components;
 - c. filtering "non component" terms, deleting rows containing words adjacent to a numeric character;
 - d. among those rows containing a same numeric character, recognising synonyms and analogue words;
 - e. identifying an intersection set of words belonging to the rows containing the same numeric character, such a set of words being assumed as a representative name of the component referenced by the numeric character of those row.
3. Method according to Claim 1, characterized in that the identification of all system components described in the examined text is performed with an assumption that said components interact as subjects and objects of a basic

functional triad TFA (Tool, a subject; Field, an action; Artifact, an object) according to the following procedure:

- a. extracting from each sentence a triad TFA (Tool-Field-Artifact), for example from an XML document or by using a semantic processor;
 - b. filtering the triads TFA containing a list of verbs not significant from a functional point of view;
 - c. collecting Tools and the Artifacts that have survived the previous filtering step;
 - d. (optionally) adding a further set of candidate components by using commonly available techniques to identify words representative of a content of a text;
 - e. among all the candidate components (Tool and Artifacts that survived the filtering phase), removing noun repetitions, also taking into account synonyms of candidate components.
4. Method according to Claim 1, characterized in that a detail/abstraction comparison criteria is applied to classify system components according to the following steps:
- a. analysing descriptive locutions and of specification's expressions like "... of ...";
 - b. assigning to a component preceding a preposition "of" a role of subsystem of a component following the same preposition "of";
 - c. searching descriptive verbs like "to comprise", "to be made of", "to be constituted by" etc., taking into account all forms that these verbs can assume, also due to conjugation irregularities;
 - d. assuming that components preceding a descriptive verb are subsystems/supersystems of components following the descriptive verb itself as function of a meaning of such a verb;
 - e. performing such an analysis taking into account all alternative denominations of each component.

5. Method according to Claim 1, characterized in that a Detail Level (DL) is assigned to each component, said DL representing a maximum abstraction level by a DL=0, each subsystem being one level greater than the DL of a corresponding supersystem.
6. Method according to Claim 5, characterized in that for several Detail Levels (DL) assigned to a same component, so that a maximum abstraction level is represented by a DL=0 and the DL of each subsystem is one level greater than the DL of a corresponding supersystem, a hierarchy simplification is performed eliminating all hierarchical jumps.
7. Method according to Claim 5, characterized in that for a same Detail Level (DL) assigned to a same component so that a maximum abstraction level is represented by a DL=0 and the DL of each subsystem is one level greater than the DL of a corresponding supersystem, a parallel hierarchy identification occurs taking into account such "parallel" hierarchies.
8. Method according to Claim 1, characterized in that all components are further processed to identify a role of a component in an assembly described in a text according to the following procedure:
 - a. an attribute "portion" is assigned to all components whose name contains words describing a portion of a component, as "end", "side", "face", "part", etc.;
 - b. an attribute "assembly" is assigned to all components having at least a subsystem that in the previous step has not been labelled as "portion";
 - c. an attribute "part" is assigned to all components not labelled in the previous two steps.
9. Method according to Claim 1, characterized in that an identification of functional links existing between recognized components of the examined system is performed according to the following steps:
 - a. searching for sequences of words containing names of two system components separated by a verb, excluding a triad

- component-verb-component so that the verb is not significant from a "functional point of view";
- b. assuming components that precede and follow said verb as the Tool and the Artifact of the triad, as function of the meaning and of the form (active/passive) of the verb itself;
 - c. searching for sequences of words containing at least one system component and a verb of the functionalities requested in a given field of application (significant verbs from a functional point of view).
 - d. assuming said component (referred to step c.) as the component of the triad, as function of the meaning and of the form (active/passive) of the verb itself.
10. Method according to Claim 9, characterized in that an external system is identified, said external system being a Tool or an Artifact of a functional triad TFA, so that it has not been recognized following criteria according to Claims 2 and 3.
 11. Method according to Claim 1, characterized in that if a functional link is identified according to the method according to Claim 9, so that the Tool is a component of the system and the pair Field-Artifact can be translated into a function, the search for the Artifact of such a function can be demanded to a user or performed by looking for a first identified component following the preposition typically associated to that pair Field-Artifact
 12. Method according to Claim 1, characterized in that all secondary products of the examined system are identified according to the following procedure:
 - a. each Artifact is assumed as a secondary product of the examined system;
 - b. a secondary product loses this property (so becoming a "standard" component of the system) in the following cases:

- in the detail level hierarchy a candidate secondary product has at least two abstraction levels above, i.e. its DL (Detail Level) is greater than or equal to 2;
 - a number of functional interactions so that a candidate secondary product is a Tool that is greater than or equal to the number of functional interactions so that it is an Artifact.
13. Method according to Claim 1, characterized in that the main product of the examined system is identified, among all identified secondary products, as one whose ratio between the number of interactions so that said secondary product is an Artifact and the number of interactions so that said secondary product is a Tool, is maximum.
14. Method according to Claim 1, characterized in that the main product of the examined system is identified among all identified secondary products, as the one whose sum of the following different probability values is maximum:
- a. checking if a secondary product is mentioned as an Artifact in the first two claims of the patent;
 - b. checking if a secondary product is mentioned as an Artifact in the abstract of the patent;
 - c. checking if a secondary product is mentioned as an Artifact in the title of the patent;
 - d. evaluating how many times the secondary products are mentioned in the whole patent and normalizing these values with respect to the maximum frequency; this normalized value multiplied by 100 is assumed as the partial probability value, but in any cases it must be lower than or equal to a predefined value;
 - e. checking if a secondary product is an Artifact of a Field present as a Field in the first two claims of the patent as well;
 - f. checking if a secondary product is an Artifact of a Field present as a Field in the abstract of the patent as well;
 - g. checking if a secondary product is an Artifact of a Field present as a Field in the title of the patent as well;

- h. evaluating how many times the Fields acting on the secondary product (considered as an Artifact) are mentioned in the whole patent and normalizing these values with respect to the maximum frequency; this normalized value multiplied by 100 is assumed as the partial probability value, but in any cases it must be lower than or equal to a predefined value;
 - i. evaluating how many times the pairs Field-Artifact, so that the Artifact is a secondary product, are mentioned in the whole patent and normalizing these values with respect to the maximum frequency; this normalized value multiplied by 100 is assumed as the partial probability value, but in any cases it must be lower than or equal to a predefined value.
15. System for performing automatic analyses and comparisons of patents and technical descriptions of engineering systems according to the method of Claim 1, said system comprising:
- a Temporary Storage Database 20 in which a text to be analysed, entered by a user, is stored;
 - a Database of Stop Words and Analogue Words 40 and (optionally) of a commercially available semantic processor (external to said system);
 - a Text Analyser Module 30 by which the text is processed;
 - a Database of Extracted Information 50;
 - a Post Processing Module 60;
- characterized in that:
- a Components Recognition module 31 allows identifying all system components described in the examined text (i.e. for a patent the components of the invention) ;
 - a Components Classification Module 32 orders and classifies the identified components;
 - an Interactions Analysis Module 33 allows identifying all functional links existing between the recognized components of the examined system;
 - all identified links are stored in the Database of Extracted information 50;

- a Product Identification Sub-Module 35 identifies all secondary products and among these a main product of the examined system;
- the Post Processing Module 60 supplies the content of the Database of Extracted Information 50 to the user, organizing such data in different forms as function of the scope of the analysis.

16. System according to Claim 15, characterized in that the identification of all system components described in the examined text is performed with a commercially available semantic processor (for example Cobrain™, Knowledgist™, Phrasys™, Semantic Explorer™, CREAX, Kiwilogic™ etc.), hence extracting from each sentence a triad TFA (Tool-Field-Artifact) through the following steps:
 - a. filtering the triads TFA (Tool-Field-Artifact) containing a Field belonging to set f) of the Stop Words and Analogue Words Database 40;
 - b. collecting the Tools and the Artifacts that have survived the previous filtering step;
 - c. (optionally) adding a further set of candidate components by using commonly available techniques to identify words representative of the content of a text (i.e. statistical analyses, cluster engine, Bayesian network etc.);
 - d. among all candidate components (Tool and Artifacts that survived the filtering phase) noun repetitions are eliminated also taking into account the synonyms list of set b) of the Stop Words and Analogue Words Database 40;
 - e. all remaining components are assumed as the components of the examined system.
17. System according to Claim 15, characterized in that the identification of the functional links existing between the recognized components of the examined system is performed with a commercially available semantic processor (for example Cobrain™, Knowledgist™, Phrasys™, Semantic Explorer™, CREAX, Kiwilogic™ etc.), hence extracting from each sentence a triad TFA (Tool-Field-Artifact) through the following steps:

- a. if both Tool and Artifact are system components and the Field is not belonging to set f) of the Stop Words and Analogue Words Database, then that TFA triad is assumed as a basic functional block of the system;
 - b. otherwise, if just one among the Tool and the Artifact is a system component, but the Field is a verb of the functionalities requested in a given field of application, then the missing Tool/Artifact is assumed as an External Component of the system and the complete triad is assumed as a basic functional block of the system;
 - c. if a pair Field-Artifact among those extracted by the semantic processor belongs to set g) of the Stop Words and Analogue Words Database 40, then the subject of the verb is assumed as the Tool of the triad and the pair Field-Artifact is translated according to set g) table of the Stop Words and Analogue Words Database 40 in a functional Field.
18. System according to Claim 15, characterized in that attributes as "assembly", "part" or "portion" identifying the role of a component in the assembly described in the text can be transferred through commonly used data exchange formats like IGES, STEP, IDEF etc.
 19. System according to Claim 15, characterized in that said attributes as "assembly", "part" or "portion" identifying the role of a component in the assembly described in the text can be linked to a geometric database of a CAD system as a direct link to a Feature Tree of a Part model and/or to an Assembly Tree of an Assembly model, hence integrating a conceptual model of a mechanical system to its embodiment.
 20. System according to Claim 15, characterized in that all identified triads, as well a position in the examined text of the sentence from where such a triad has been extracted, are stored in the database of Extracted Information 50.
 21. System according to Claim 15, characterized in that a position in the examined text of the sentence from where such a triad has been extracted, is evaluated just numbering with a sequential order all sentences of the examined text,

- distinguishing a sentence from another on the basis of the character "." or the ASCII character Carriage Return.
22. Post Processing module 60 of a system according to Claim 15, characterized in that a Text Content Module 61 represents:
- a. each identified component of the system with its reference number and the representative name defined by the Components Recognition Module 31;
 - b. each identified component or subject external to the system;
 - c. the main product for internal/external components;
 - d. the detail level hierarchy determined by the Classification Module 32;
 - e. the functional interactions between the identified components according to the results of the Interactions Analysis Module 33.
23. Text Content Module 61 of a system according to Claim 22, characterized in that it represents:
- each identified component of the system by a rectangle labelled with its reference number and the representative name defined by the Components Recognition Module 31;
 - each identified component or subject external to the system is represented by an hexagon labelled with the string "EXT";
 - a sequential number and the representative name defined by the Components Recognition Module 31;
 - the main product by an ellipse labelled with the same criteria illustrated above for internal/external components;
 - the detail level hierarchy determined by the Classification Module 32 represented nesting the components at a deeper detail level inside the components at a more abstract level;
 - the functional interactions between the identified components is represented with arrows pointing from the Tool to the Artifact, labelled with the Field, according to the results of the Interactions Analysis Module 33.
24. Post Processing Module 60 of a system according to Claim 22, characterized in that said Text Content Module 61 represents:

- a. a list of components with their detail level DL and a corresponding supersystem;
- b. a list of secondary products as pairs Field-Artifact with their main product probability value MPPV;
- c. a list of partial probability values evaluated according to the procedure detailed in the description of the Products Identification Sub-Module 35;
- d. a list of functional interactions between the identified components.

25. Post Processing Module 60 of a system according to Claim 22, characterized in that a Text Comparison Module 62 allows the comparison between two or more systems descriptions according to the following parameters:

- a. comparison between a "system diameter", that is a number of detail levels identified by the Components Classification Module 32;
- b. comparison between a number of internal components of the examined systems, both taking into account the whole list of identified components and each detail level;
- c. (if the analysis of the Mechanical Embodiment Analysis Sub-Module 36 has been performed) comparison between a number of "assembly", "part" and "portion" of the examined systems;
- d. comparison between a number of interactions identified by the Interactions Analysis Module 33; if two or more Fields are associated to the same pair Tool/Artifact a check to eliminate synonymous Fields is performed taking into account the set i) of the Stop Words and Analogue Words Database 40;
- e. comparison between a number of interactions (counted as in step d) acting on components at a same Detail Level; it can be highlighted if these components belong to the same supersystem or not;
- f. comparison between a number of interactions (counted as in step d) acting on components at a different Detail Level; it can be highlighted if these components are one

- subsystem of the other or not; it is also taken into account the "hierarchical distance" between the interacting components, i.e. the difference between their detail levels;
- g. comparison between number and lengths of branches being present in the functional diagram of the examined systems (as the one in Fig. 6) evaluated starting from a Main Product of the systems themselves;
 - h. comparison of components at a same rank: the rank of a component is defined as a minimum distance, in terms of number of interactions, that links the Main Product of the system with the component itself;
 - i. analysis of a detail level run along the description of the examined system: the Interactions Analysis Module 33 stores a position in the text of each identified interaction TFA; the detail level of the Tool and the Artifact in a sentence is assumed as the detail level of the description, hence it is possible to analyse the detail level run in the examined text and to compare such a run in different texts.
26. Post Processing Module 60 of a system according to Claim 22, characterized in that the analysis of the peaks of the Detail Level runs along the description of a system allowing the identification of the core and the secondary peculiarities of the system itself.
27. Post Processing Module 60 of a system according to Claim 22, characterized in that a Database of Functional Usage of Components in Different Systems 63 stores all functional interactions associated to homonymous components in all examined texts, recording a reference to a source text and the role of the component in the TFA triad.
28. Post Processing Module 60 of a system according to Claim 22, characterized in that a Database of Components Capable of Performing a Given Function 64 stores:

- a. all Tools associated with homonymous Fields found in all examined texts, recording a reference to a source text and a complete TFA triad;
 - b. all Tools associated with homonymous pairs Fields-Artifacts found in all examined texts, recording the reference to the source text and the complete TFA triad.
29. Stop Words and Analogue Words Database 40 of a system according to Claim 15, characterized in that it is constituted by eight set of words, all customisable by a user through the following automatic extraction procedure:
- a) the user supplies to the system a set of typical documents of the field of application he is interested in;
 - b) a semantic analysis is performed through a commercially available semantic processor (for example Cobrain™, Knowledgist™, Phrasys™, Semantic Explorer™, Kiwilogic™ etc.) and a table of Tools and Artifacts and their occurrence is stored;
 - c) by comparing the table defined in the previous step and the complete Database 40 it is possible to customise automatically the Filters and Synonyms lists, hence creating typical subsets of the Database 40 labelled with the field of application of the documents processed at the step 1).
30. Stop Words and Analogue Words Database 40 of a system according to Claim 29, characterized in that said database is constituted by the following sets:
- a) list of stop keywords for words adjacent to numeric characters during the Components Recognition task; this set is typically constituted by references to Figures, Patents or other documents, units etc.
 - b) table of synonyms of candidate components, at different detail level (for example, portion, side, end; piston, plunger etc.);
 - c) list of typical Fields of the functionalities requested in a given field of application;

- d) table of descriptive verbs like "to comprise", "to be made of", "to be constituted by", etc., such a list having to take into account all forms that these verbs can assume, also due to conjugation irregularities;
- e) list of terms describing a portion of a component, as "end", "side", "face", "part" etc.
- f) list of verbs not significant from a functional point of view;
- g) g) table of the pairs Field-Artifact, their translations in a functional verb and one or more prepositions typically associated to that locution, used to search the Artifact automatically;
- h) table of synonyms of functional verbs representing a Field.

31. System according to Claim 15, characterized in that the customisations of the following systems are allowed:

- a. activities of Components Recognition Module 31, Component Classification Module 32 and Interactions Analysis Module 33 can be followed step by step by the user, who may compare the extracted information with its source sentence, or can be performed automatically even if with a lower reliability;
- b. the user can specify a list of components (Tools/Artifacts) and/or functions (Fields) to focus the Interactions Analysis on, so that just the corresponding functional sub-diagrams are extracted;
- c. the search for Secondary Products and/or the Main product of the examined systems can be limited to the components external to those systems.